INDIGENOUS LANGUAGE TECHNOLOGIES: ANTI-COLONIAL OASES IN A COLONIZING (DIGITAL) WORLD

Nathan Thanyehténhas Brinklow [1]
Queen's University, Kingston, ON, Canada
Tsi Tyónnheht Onkwawén:na, Tyendinaga Mohawk Territory

Abstract

Language technologies are an increasingly common part of daily life for people around the world. Millions of users per day access services like Google Translate and Apple's Siri in a technology ecosystem that favours a handful of the world's most common languages. In a form of digital colonization, Indigenous languages are pushed aside in a profit-based system of research and development that results in both values conflicts and technological misalignments. Despite the hostile environment that disincentivizes the use of Indigenous languages, Indigenous language communities are pushing back by engaging language technologies to proactively support their work of language maintenance and revitalization. This paper argues that Indigenous leadership in the development of language technologies encourages the development of responsive and responsible Indigenous language technologies (ILT) that push back against dominant cultural and technical limitations.

Keywords: Indigenous language technology, decolonizing technology, digital colonization, speech technology

Citation: Brinklow, N. (2021). Indigenous language technologies: Anti-colonial oases in a colonizing (digital) world. *WINHEC: International Journal of Indigenous Education Scholarship*, *16*(1), pp. 239-266. http://dx.doi.org/10.18357/wj1202120288. Special Issue on *Indigenous Language Revitalization: Innovation, Reflection and Future Directions,* Guest Co-Editors Drs. Onowa McIvor and Kari A. B. Chew.

[1] Correspondence: Nathan Thanyehténhas Brinklow, Queen's University, nathan.brinklow@queensu.ca

Introduction

Technological engagement and innovation are critical components for the ongoing survival of Indigenous languages around the world. In the North American context, the shift to electronically mediated communication has been underway for at least a generation (Buszard-Welcher, 2018) and the COVID-19 pandemic is further accelerating this trend with the health risks associated with in-person communication (Chew, 2021; Sinclair, 2020). In this rapidly changing environment, Indigenous languages are at risk of being left behind in the digital space by technological changes driven by large multinational corporations supporting a small handful of the world's most common languages. The spaces and tools created by these large companies impose a set of values about languages and impose languages themselves at the expense of minoritized and Indigenous communities and their languages. This ubiquity of these tools, the monopolies of these corporations, and the imposition of language values represent an ongoing process of digital colonization. In this exploitative context of businesses and languages, it is critically important for Indigenous people to take active roles in the development of responsive and responsible Indigenous language technologies (ILT) that prioritize the voice and values of their respective language communities (Galla, 2016; Lothian et al., 2019).

For the purpose of this paper, Indigenous leadership is defined in the most basic sense, that is, Indigenous people and speakers of Indigenous languages involved in all aspects of technology development, from identifying the problems, finding solutions, and exercising control, to deploying the tools back to their communities. The ideal of Indigenous leadership in the development in ILTs faces challenges that are shared with other minority language

communities. Even as the tools for creating advanced language technologies become easier to access, they remain expensive and time consuming to develop, typically requiring large amounts of language data which are not always available in Indigenous and minority contexts. Despite a challenging environment for the development of ILTs, Indigenous participation and leadership in language technology development actively challenges dominant technological and value assumptions and creates anti-colonial oases where Indigenous and minority languages can flourish.

Background

Adapting to new language technologies for communication is nothing new, and history is replete with examples of shifting language technologies. In their simplest form, language technologies are the tools and methods used for encoding knowledge and communication, including everything from signal fires to satellite communications. Individual language technologies come and go as new tools arise and others fall out of use, while change remains the constant. This perpetual change is seen in the evolution of writing instruments from stones and charcoal, to brushes, quill pens and pencils, to mechanical typewriters and then finally computer keyboards. Each of these tools was a response to a new context that demanded a new solution as human needs changed. These shifts came about through advances that enabled new and more efficient means of communication, paralleling the development of technology in general. The modern experience of global communication can find its roots in the ships that carried messages around the world in ancient times and gave way to telegraph wires laid under the oceans and around the world throughout the second half of the 19th century. The global scope of communication technology was further amplified

with the creation of the first digital computers in the 1940s that brought language technology into the present digital age.

The digital transition of language technologies has continued unabated since the invention of the digital computer, and modern language technologies blend into the background of daily life and work for people around the world (Buszard-Welcher, 2018). Ubiquitous technologies like mobile telephones and text messaging; the Internet and social media; spelling and grammar checkers; automatic translation; and speech technologies like "Siri" and "Alexa" enable continuous and effortless communication across the room and around the world. While this interaction with technology is routine for many, it is mediated through a small number of the world's most common languages and often provided by a small number of multinational corporations. This limited interface with unlimited potential is of critical importance to Indigenous language communities that want to have a digital presence.

Just Another Colonizer? Technology and Language Shift

Majority languages place enormous pressure on lesser-spoken languages as speakers tend to migrate, willingly or otherwise, to languages with a (perceived) higher socio-economic value. This is a basic fact of language shift and was a common tool in the assimilation of Indigenous people in Canada (Truth and Reconciliation Commission of Canada, 2015). The pressure to shift languages is amplified in the digital sphere where the world's major languages hyper-dominate available technologies and services. While the actors have changed, the imposition of particular languages and underlying values about language represent an active process of colonization. Before the widespread adoption of digital

language technologies, Bernard reported that "about 97% of the world's people speak about 4% of the world's languages; and conversely, about 96% of the world's languages are spoken by about 3% of the world's people" (1997, p. 142).

The linguistic hegemony of the world's major languages is multiplied online where large companies drive technological development and content creation for these languages. As an example, YouTube reports that their service is available in 80 languages reaching 95 percent of the world's online population (YouTube About, n.d.). This statistic from YouTube, while inherently skewed toward wealthier countries and people with Internet access, clearly demonstrates the negative economic pressure facing development of technologies for the vast majority of the world's (non-majority) languages. With the ability to serve so many people with so few languages, there is no financial incentive for companies like YouTube, Google, Microsoft, Apple, Amazon to expand their services to include any of the thousands of other languages around the world.

As with language technologies, languages themselves change over time. These changes occur through alterations of the socio-political climate, migration and interaction with neighbouring peoples, through natural processes of language evolution, or through adoption of lingua franca for shared communication in specific language domains. The natural expression of a multilingual environment is now challenged by rapidly changing technologies, and change is hastened by instantaneous global communication. In this new context, language choice is no longer limited by time, geography, or socio-politico boundaries, but rather it is *imposed* by the same technology that enables communication.

The factors contributing to language shift in the current situation are complex, and they are being explored; however, it is impossible to ignore the fact that technology itself may be a significant factor in the language pressures faced by Indigenous language communities (Galla, 2016). The threat posed by the pervasiveness of the language technologies of world's major languages creates a complex relationship between Indigenous communities and the language technologies they see as vital to preserving their languages into the future. Galla (2018) conveys the complexity of this relationship as the proverbial "two-edged sword," recognizing the harm inflicted on Indigenous communities by hegemonic languages, but also feeling the necessity of active engagement with technology to ensure language survival.

Many Indigenous language communities are looking at ILTs as potential tools for Indigenous language revitalization (ILR) (Galla, 2009; Keegan & Cunliffe, 2014; Kuhn et al., 2020; Littell et al., 2018; Wagner, 2017), and it is exciting to think of talking to "Siri" in a small Indigenous language. However, with the negative financial pressure facing the development of ILTs by the world's large technology companies, Indigenous people must take the lead in developing the next wave of responsive and responsible language technologies. With the increasing importance and use of language technologies for all languages, it is more important than ever for Indigenous language communities to carry their languages through the digital transition to ensure the languages' continued use and vitality.

Moving into Colonized Spaces (Digital Transitions)

As communication and activities of daily life continue their transition into the digital sphere, Indigenous people are finding and creating spaces for their languages in the mainstream

digital world (Keegan & Sciascia, 2018; Littell et al., 2018; McIvor & Anisman, 2018). The idea of creating space is nothing new for Indigenous peoples who have existed for generations in a hostile milieu that transcends land, politics, language, resources, and religion. The digital spaces being created today differ according to the needs and capacity of communities and their languages, but they all share the goal of normalizing, stabilizing, and revitalizing Indigenous languages.

Some technologies of daily life are easier to adapt than others. An easily accessible space for many language communities is social media. Green (2017) highlights the role of social media engagement as a strategy for advanced language acquisition in adult Kanyen'kéha (Mohawk) language learners at Six Nations, in Ontario, Canada. Likewise, other language communities are engaging social media as a language teaching tool (Blake, 2017; Outakoski et al., 2018). The politics and values of social media spaces like Facebook or Twitter may not always align with Indigenous values, but these platforms are freely available and widely used. However, these sites are again dominated by the world's major languages, and Indigenous language users must typically navigate the site in a majority language.

The reality of digital technologies dominated by a few languages highlights the need for the creation of Indigenous spaces in the digital transition. In contexts with large numbers of speakers, such as Aotearoa (New Zealand), this transition is enabling the use of Indigenous languages in the activities of daily life, for example using a bank machine or checking out a book from the library (Keegan & Sciascia, 2018), where a translation layer sits on top of a common technology. The spaces occupied by these speaking communities enable Indigenous

languages to live on in modern life and normalize the presence of Indigenous languages in the digital world. Keegan and Cunliffe (2014) state that passing on te reo Māori (the Māori language) to the next generation is insufficient to ensuring language survival if there are no opportunities to use the language. This attitude inspires ongoing technological development and helps ensure that te reo Māori can be used in daily life (Keegan & Nfato, 2014; Mato et al., 2016; Whaanga et al., 2015).

For languages with fewer speakers, digital transitions are assisting language revitalization through computer assisted language learning (Bontogon et al., 2018; Kazantseva et al., 2018; Lessard et al., 2018). In the Canadian context, this approach is responsive to the needs of language communities involved in language revitalization through structured language teaching and acquisition programs. These tools are typically designed to help learners acquire the language and explore features of the language while creating opportunities for learners to boost their language proficiency away from the classroom. A prominent example of this technology is firstvoices.com, which "is a suite of web-based tools and services designed to support Indigenous people engaged in language archiving, language teaching and culture revitalization" (First Peoples' Cultural Council, n.d., para. 1). FirstVoices provides a standard set of language independent (or agnostic) tools that are essentially containers that hold whatever language data is placed into them. Digital technologies are also supporting Indigenous languages by enabling language documentation that drives both language programming and the development of advanced language technologies (Bird, 2018; Rice & Thieberger, 2018).

It's Not Easy Living in Colonized Spaces (Technical Challenges)

Indigenous communities experience varying levels of difficulty in completing this digital transition. [2] Looking back to one of the foundational technologies discussed earlier, orthography can be a major barrier to being online for Indigenous languages, meaning the ability or inability to accurately represent Indigenous writing systems. The orthographic barrier is one more colonizing factor that is critically important to overcome as text is a primary medium of interaction with the digital world.

As defined by Schillo and Turin (2020, p. 72), "orthography refers to writing conventions that are implemented when using a script, such as capitalization, or the set of letters from a script used by a particular language." In a digital environment, the script or writing system can be the orthographic barrier. For languages that use the standard Latin or Roman characters printed on this page, a digital transition is simplified as digital environments support the letters, even if they do not support the language. The distinction is important because it allows a language like Kanyen'kéha (Mohawk), which uses the standard Latin/Roman alphabet with diacritics common in European languages (The Mohawk Language Steering Committee, 1993), to be used in digital environments, even without language-specific support with tools like spell checkers or grammar aids.

A challenge exists for languages that use their own writing systems. In the North American context, the ᏣᎳᎩ (Cherokee) syllabary developed by Sequoyah in the early 19th century is

[2] For an overview of the process of digitization, see *Indigenous Languages: Zero to Digital* (2019). The non-profit group Translation Commons created this guidebook in partnership with UNESCO as part of the 2019 International Year of Indigenous Languages.

an historically significant Indigenous script. This writing system uses 85 individual characters to represent whole syllables of the language rather than individual phonemes or letters (King, 1975, pp. 11–12). In the 19th century, the syllabary allowed the ᏣᎳᎩ to claim space in the print world through letterpress printing (another language technology innovation). The transition to written language opened new domains for expression, and Cherokee literacy surpassed that of the settler population (Parins, 2013). The syllabary was later adapted to typewriters and then made a digital transition with the design of ᏣᎳᎩ keyboards and typefaces (fonts). This process mirrors that of languages in Canada like Cree and Inuktitut that also use syllabic systems, and it illustrates that orthographic barriers in technology can be overcome, although with some difficulty (Schillo & Turin, 2020). Unfortunately, the availability of a typeface or font does not guarantee universal access across multiple devices and programs that need to be programmed to recognize additional characters and diacritics.

The problem of universal readability is addressed through Unicode.[3] Unicode is a global standard that ensures characters display correctly across multiple devices, sites, and programs. When a character is included in Unicode, any compliant technology should accept and display that character correctly. If characters are not Unicode compliant, they may not display correctly and/or force users to use other alphabets. Communities that use their own writing systems can apply to have their characters included in the global Unicode standard, but this presents a significant technical challenge. In their discussion of new technologies for

[3] For more information, see https://home.unicode.org

Indigenous languages, Buszard-Welcher (2018) asserts the importance of Unicode compliance as a foundational technology for digital transition. Likewise, The Unicode Consortium itself declares on their homepage that "everyone in the world should be able to use their own language on phones and computers" (The Unicode Consortium, n.d.). However, the process of proposing new characters for the global standard is highly technical and requires expertise which can be beyond the capability of many under-resourced language communities. The complexity of Unicode adoption is illustrated by Pine and Turin (2018) who oversaw the modernization and adoption of Haíłzaqv (Heiltsuk) characters into the global Unicode standard. As of March 2020, Unicode 13.0 includes 143,859 characters, including emoji (The Unicode Consortium, n.d.). The process of Unicode adoption remains overly complicated and is only a first step in bringing an Indigenous language into the digital sphere.

On the surface, services offered by Facebook or Google permit the use of Indigenous languages; however, the underlying language technologies privilege the world's major languages. This foundational dependence on the major languages is clear when one considers that Google's search algorithms must "understand" a language to suggest related topics and "Alexa" must speak the language of a user. At most, the technologies that power these services are anchored in a few hundred of the world's more than 7000 languages. Adding to this concern, the users of these services ultimately depend on the goodwill and continued profitability of large multinational companies to ensure continued access.

Essentially, Indigenous people are claiming digital spaces for their languages in a hostile environment. Google's self-declared aim to "organize the world's information" (Google, n.d. para. 3) necessitates language technologies that categorize, store, translate, and provide this information to users around the world. This process can only happen through majority languages, and the use of services provided by the large technology companies is essentially the passive use of inherently non-Indigenous spaces. The users (Indigenous or otherwise) of these services do not have a direct role in the development or design of the services provided by these companies and have little control over the way their information is used. The services provided by these companies are directed at a global market of consumers where overall value is determined by the largest possible return for the smallest possible investment. This defining value presents a conflict for many Indigenous communities.

The New Colonizers (Systemic Challenges)

The world's top five technology companies (Apple, Google, Amazon, Facebook, and Microsoft) drive the development of language technology for the major world languages in a profit-based model. It may not always be clear to the user, but each of these companies is selling something to someone, even when providing "free" services. The profit-based model works well for majority languages because it supports the creation of language technologies that are expensive and time consuming to develop. However, the profit-driven model for language technology and content development is disadvantageous for Indigenous peoples and their languages as it favours a handful of widely spoken languages at the expense of thousands of smaller (and often Indigenous) languages.

At the LT4ALL (Language Technology for All) gathering hosted by UNESCO (Paris, December 2019), Frances Tyers characterized the development of mainstream language technologies as focused on "small numbers of rich people, or large numbers of poor people." Tyers' critique was pointed squarely at the large technology companies and their profit-driven motivation. In Tyers' four-part division (rich/poor, many/few), most Indigenous language communities fall in the underserved intersection of small numbers of poor people. While there are occasional forays into the world of endangered languages by the large technology companies,[4] the companies are ultimately responsible for delivering profits to shareholders, and any other activities are (at best) ancillary to their central mission. The global dominance of a few companies and their often proprietary language technologies creates a digital space focused on a small number languages that are accessible at the expense of all others.

The dominance of the world's major languages across the Internet is sustained by the underlying language technologies that are largely designed for English. In these cases, even major world languages face challenges adapting these technologies to their languages. Gilles Boulianne describes his experience working with speech technologies for Canadian French as one of constantly finding ways to adapt inherently English technologies to French needs (personal communication, November 19, 2019). This challenge is further amplified when adapting these underlying English technologies to completely unrelated Indigenous languages. While technologies like artificial intelligence are theoretically language agnostic,

[4] For example, Google oversaw the development of the Endangered Languages Project (http://www.endangeredlanguages.com/about/) before turning over ownership to expert organizations, and Microsoft has localized some of their software packages into a number of Indigenous languages including Māori, Welsh, Catalan, and Cherokee with translation support from those language communities.

the data requirements are an effective barrier to all but a handful well-resourced Indigenous languages (Buszard-Welcher, 2018).

In their assessment of challenges to ILT in Canada, Littell et al. (2018) outline a number of projects that are attempting to adapt these technologies, including text-to-speech and automatic speech recognition (ASR) for Indigenous languages. Advances in computing power, programming, and algorithm design are creating opportunities for communities to adapt these advanced technologies (Jimerson & Prud'hommeaux, 2018; Lessard et al., 2018; Micher, 2018), but work is ongoing and their ultimate accuracy and effectiveness is unknown at this point.

The profit-based model that drives the global technology giants also devalues non-majority languages when it prioritizes provision of service to the most people at the lowest cost. By default, this model devalues ancestral and heritage languages and relegates them to non-digital spaces, diminishing their value for the future of electronically mediated communication and contributing to language shift. This profit-first valuation of language is the polar opposite of work by countless Indigenous language communities to restore and revitalize their languages for their inherent value. Indigenous leadership in the development of ILT challenges the system as it exists and has the potential to create technologies that support Indigenous languages rather than continue their devitalization. Fortunately, Indigenous people around the world are exercising technological leadership as they claim spaces for their languages in the digital realm.

A Note About Data Sovereignty

Given the hostile environment created by a profit-based multinational model, responsive and responsible language technology development must be done with care and planning to protect Indigenous languages from the effects of digital colonization and potential exploitation. In the context of an online existence, Indigenous communities around the world are increasingly thinking about the role of data sovereignty in their digital futures (Davis, 2016; Pool, 2016; Walter & Suina, 2019; Wilks et al., 2018). Data sovereignty is briefly defined as Indigenous people maintaining control over their own data and digital development. In the realm of ILT, language data includes parallel translations, computer codes that interpret or construct language, texts in Indigenous languages, and voice recordings. Responsible language technology design ensures that language communities maintain control over their language data at all stages of technological development.

Indigenous leadership within ILT development keeps the voice of the language communities at the forefront and can help ensure that the underlying values of the language technologies are rooted in Indigenous values. A discussion of values vis-à-vis technology may seem out of place given the common narrative of value-neutral technologies; however, as Miller (2020) argues, technology is part of the socio-politico-cultural system, not apart from the system. Data sovereignty (and community sovereignty) are values questions that need to be considered in ILT development along with value-laden questions such as language ownership, who has the right to learn a language, and how languages should be protected. Those questions are outside the scope of this paper but illustrate the complexity of creating anti-colonial spaces for Indigenous languages to thrive in a digital environment.

In his keynote address to the HELISET TŦE SḰÁL conference (June 2019), Keegan shares his (negative) experience of working with Microsoft and Google to develop resources for te reo Māori. The model used by these two projects alienated language data from the community that created it and absorbed the data into various software platforms and algorithms. Keegan shares that the community was not able to keep its own translations and encourages Indigenous communities to ensure control of their language data. With his experience, Keegan asserts that "we're the only ones that really *care* about our languages, so if we want something, we have to make it ourselves" (Keegan, 2019). While Dr Keegan's encouragement to ensure "we do for ourselves" is certainly the ideal way forward, many language communities lack capacity to undertake this technically challenging work that combines linguistics, computer science, and artificial intelligence with Indigenous knowledge.

For Indigenous languages around the world, the process of technological shift has created new mediums of expression. As Galla notes, "digital technology has created new domains for languages to exist, allowing learners and speakers to engage in or at least feel that the language is a necessary part of their Indigenous well-being and the contemporary world" (2016, p. 1123). New technologies present new opportunities to ensure the continued survival of Indigenous languages.

Anti-Colonial Oases

To meet the challenges outlined above, many communities in Canada are collaborating with non-profit partners for the development of ILT outside of a profit-driven ecosystem (Kuhn

et al., 2020; Littell et al., 2018; Rice & Thieberger, 2018). These innovative partnerships include universities, communities, research institutions, governments, and others, which go beyond simply decolonizing these spaces into active anti-colonial work that challenges systemic and technological limits to advance the presence of Indigenous languages in the digital world.

In the Canadian context, these partnerships (ironically, often funded by the original colonizing governments) are producing ground-breaking language technologies that are innovative by any standard. One pioneering partnership has developed between Onkwawenna Kentyohkhwa at Six Nations in Ontario and the National Research Council (NRC) of Canada. The partnership is between the Indigenous Language Technology Project Team and local teachers to develop a verb-conjugator for Kanyen'kéha (Mohawk) (Kazantseva et al., 2018). This application emerges from the community's desire to support learners in their exploration of complex Kanyen'kéha verb morphology. Kawennón:nis (the Word Maker) is the product of technically challenging work, with the community providing language expertise and the NRC programming skill and financial resources. Furthermore, while the initial development was done with a specific language (Mohawk), the underlying tool was created to work with any language in a "first deep, then broad" approach to design and development (National Research Council of Canada, n.d.). The partnership with the NRC adds value to the project at the national level because a broad approach to development is not the responsibility of an individual language community.

In the realm of speech technology, the simplicity of interactions with "Siri" or Google Translate disguises the complexity of the underlying technologies. These technologies are in demand (Littell et al., 2018) but are very difficult to develop. However, the increasing accessibility of computing power and innovative programming is enabling research and development that would ordinarily be limited to high-resource languages (Besacier et al., 2014; Gupta & Boulianne, 2020a; Jimerson et al., 2018). Speech technologies have great potential to support language communities and language learners through tools such as voice recognition and talking dictionaries, along with tools for language documentation (Cox et al., 2019; Zahrer et al., 2020).

Jimerson and Prud'hommeaux (2018) report on their work for ASR in Seneca in the Seneca homeland in what is now New York State. This project is directed by a Seneca citizen to support the nation's language learners and strengthen their language for the future. Indigenous leadership ensures that the research keeps the needs of the community at the centre and aligns with Keegan's assertion (2019) that "we do for ourselves." This innovative research for a small Indigenous nation with few speakers has great potential to support the community's language goals but would never be undertaken by a large technology company. As such, a research partnership between the language community and a university (in this case through a PhD student) is an ideal vehicle to complete this work. This type of innovative research is stimulating and challenging and there are many non-Indigenous researchers/allies who are intensely interested in helping communities solve technical challenges (Kuhn et al., 2020).

Indigenous leadership in technology development also helps address concerns over data sovereignty. As Herbert discusses in *The Financial Post* (Malone, 2017), when Indigenous people take the lead in technological development, it helps ensure that language data and intellectual property remain with the Indigenous communities. The leadership role within research partnerships needs to be one of equals to avoid the experience Keegan relates (2019) of working with Google and Microsoft and the alienation of language data from the community.

The alienation described by Keegan (2019) is rooted in a conflict of values. In the profit-driven eco-system of large technology players, the value of language data comes from the ability to offer services that will attract users and generate revenue. For Indigenous communities, especially those engaged in language maintenance and revitalization efforts, the value of language data is relational and comes from its ability to support and enhance language use and learning. In the context of these Indigenous communities, language data is "valuable" because the language itself is inherently valuable for diverse and dynamic socio-politico-cultural reasons. While there are potentially valuable opportunities to partner with large technology companies to access their platforms or technologies, the values conflict needs to be acknowledged, and community interests in their data need to be protected. Indigenous leadership in ILT development can ensure that these questions of data sovereignty are addressed in planning, development, and distribution stages.

In the mainstream technological world, software and licensing models are used to protect content creators and data while granting users varying degrees of access to programs and

services. There are two levels to this protection: the software and the data that was used to create it. In the context of language technologies, there are varying instances of closed source and open source programs, with varying levels of access to the data that are used to create language models. For the big tech companies, their programs and data sets are generally closed to protect their financial value. When the focus shifts to Indigenous led development of ILTs by communities and people working to strengthen their own languages rather than exploit their monetary value, the open/closed source models of programs and data sets becomes complicated.

Open source programs have obvious value to prevent reduplication of effort in low resource contexts (Brinklow et al., 2019). As language technology tools are developed, it is in the collective interest to see that they are shared and adaptable for other languages. This open source approach guides the work of the National Research Council of Canada Indigenous Languages Technology unit in an empowerment based approach (Kuhn et al., 2020). As tools are created in partnership with Indigenous communities, they are released for other communities and languages to adapt to local needs.

In terms of language data, for under-resourced Indigenous languages from communities with histories of exploitation, opening-up to outsiders can be a rightful cause for concern. While open source tools, free of language specific data, present a smaller threat to community data sovereignty, language data can be sensitive, depending on the priorities and beliefs of the individual communities. The models of open and closed source programs and data presume that an individual or company owns the information. When working with an

Indigenous language, the concept of licensing is stretched, with conflicting views of language ownership (Hutton, 2010) in a context based on protecting property. Indigenous values can be infused into licensing models to open programs and data for use and development while taking protection into their own hands rather than relying on a set of accepted licence models.

One such Indigenous approach to data sovereignty that can apply to programs and data sets in language technology is the Kaitiakatanga licence developed by Te Hiku Media (2018). This licensing model is based on the Māori concept of guardianship or protection and intends to create resources that are both open and protected. The application of an Indigenous licensing model can respect principles of open access while maintaining data sovereignty. While this model was developed by a solely Indigenous led and run project, without external partnerships, there are principles that are useful in partnership or sponsorship relationships.

Conclusion

Technological engagement is a critical component of language revitalization strategies for Indigenous communities around the world. This engagement takes many forms depending on the needs and desires of the various language communities and falls across the entire spectrum from low- to highly-advanced technologies (Zhao, 2003). Some communities are taking their first steps by creating digital fonts with their local orthographies (Pine & Turin, 2018) while others are conducting research that is pushing the boundaries of computer-speech interaction (Gupta & Boulianne, 2020a, 2020b; Jimerson & Prud'hommeaux, 2018). On the whole, Indigenous and minority language communities are engaging technology to

proactively support their efforts of language maintenance and revitalization (Galla, 2016; McIvor & Anisman, 2018; Ward, 2004).

Indigenous languages around the world are transitioning into the digital world as part of their ongoing process of adaptation to novel language technologies. This transition is complicated by real concerns about the role of technology as a potential cause of language shift (UNESCO Ad Hoc Expert Group on Endangered Languages, 2003). Despite the risk, many communities have concluded that a presence in the digital world will ensure the vitality of their languages for the next generation (Galla, 2016).

A digital transition brings Indigenous languages into a hostile environment that privileges a small number of common languages in a profit-driven pursuit to provide service to the most people at the least possible cost. It is only when Indigenous people take active roles in the development of language technologies *outside* of the profit-driven ecosystem, that they can help ensure the responsible development of responsive and responsible language technologies that challenge the dominant systemic and technical limitations and create anti-colonial oases in a colonizing (digital) world.

About the Author

Thanyehténhas ní:'i yónkyats. Kanyen'kehá:ka niwakhwenstyò:ten táhnon wakenyáhten. Kenhtè:ke nitiwaké:non táhnon eh nón:we kherihonnyén:nis onkwawén:na. Ó:nen oyé:ri niyohserá:ke wakateweyenstonhátye táhnon shé:kon sótsi é:so tká:yen akeweyentéhta'ne'. Tekhenonhwerá:tons ne yonkerihonnyén:nis tsi wa'onkyé:nawase' ohstónha aká:ronke' nonkwawén:na. Nyawenhkó:wa. Sewayo'tenhserí:yos.

Nathan Brinklow is Turtle Clan Mohawk from the Tyendinaga Mohawk Territory on the north shore of Lake Ontario. He teaches Kanyen'kéha for Queen's University in Kingston, Ontario and works with Tsi Tyónnheht Onkwawén:na Language and Cultural Centre at Tyendinaga. He is an adult learner of Mohawk and PhD student at the University of Victoria.

References

Bernard, H. R. (1997). Language preservation and publishing. In N. H. Hornberger (Ed.), *Indigenous literacies in the Americas: Language planning from the bottom up* (pp. 139–156). De Gruyter. https://doi.org/10.1515/9783110814798.139

Besacier, L., Barnard, E., Karpov, A., & Schultz, T. (2014). Automatic speech recognition for under-resourced languages: A survey. *Speech Communication*, *56*(1), 85–100. https://doi.org/10.1016/j.specom.2013.07.008

Bird, S. (2018). Designing mobile applications for endangered languages. In K. L. Rehg & L. Campbell (Eds.), *The Oxford handbook of endangered languages*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190610029.013.40

Blake, R. J. (2017). Technologies for teaching and learning L2 speaking. In C. A. Chapelle & S. Sauro (Eds.), *The handbook of technology and second language teaching and learning* (pp. 107–117). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118914069.ch8

Bontogon, M., Arppe, A., Antonsen, L., Thunder, D., & Lachler, J. (2018). Intelligent computer assisted language learning (ICALL) for Nêhiyawêwin: An in-depth user-experience evaluation. *Canadian Modern Language Review*, *74*(3), 337–362. https://doi.org/10.3138/cmlr.4054

Brinklow, N. T., Littell, P., Lothian, D., Pine, A., & Souter, H. (2019). Indigenous language technologies and language reclamation in Canada. In G. Adda, K. Choukri, I. Kasinskaite-Buddeberg, J. Mariani, H. Mazo, & S. Sakriani (Eds.), *Collection of research papers of the 1st international conference on language technologies for all* (pp. 402–406). European Language Resources Association (ELRA). https://lt4all.elra.info/proceedings/lt4all2019/

Buszard-Welcher, L. (2018). New media for endangered languages. In K. L. Rehg & L. Campbell (Eds.), *The Oxford handbook of endangered languages*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190610029.013.24

Chew, K. A. B. (2021). #KeepOurLanguagesStrong: Indigenous language revitalization on social media during the early COVID-19 Pandemic. *Language Documentation & Conservation*, *15*(July 2020), 239–266.

Cox, C., Boulianne, G., & Alam, J. (2019). *Taking aim at the transcription bottleneck: Integrating speech technology into language documentation and conservation*. [Conference Presentation]. International Conference on Language Documentation & Conservation, University of Hawai'i, Mānoa, HI, United States. https://scholarspace.manoa.hawaii.edu/bitstream/10125/44841/44841.pdf

Davis, M. (2016). Data and the United Nations declaration on the rights of Indigenous Peoples. In T. Kukutai & J. Taylor (Eds.), *Indigenous data sovereignty: Toward an agenda* (Vol. 38, pp. 25–38). ANU Press. http://www.jstor.org/stable/j.ctt1q1crgf.9

First Peoples' Cultural Council. (n.d.). *FirstVoices*. Retrieved December 13, 2019, from https://www.firstvoices.com

Galla, C. (2016). Indigenous language revitalization, promotion, and education: Function of digital technology. *Computer Assisted Language Learning*, *29*(7), 1137–1151. https://doi.org/10.1080/09588221.2016.1166137

Galla, C. K. (2009). Indigenous language revitalization and technology from traditional to contemporary domains. In J. Reyhner & L. Lockard (Eds.), *Indigenous language revitalization: Encouragement, guidance & lessons learned* (pp. 167–182). Northern Arizona University.

Galla, C. K. (2018). Digital realities of Indigenous language revitalization: A look at Hawaiian language technology in the modern world. *Language and Literacy*, *20*(3), 100–120. https://doi.org/10.20360/langandlit29412

Google. (n.d.). *Our mission*. Retrieved December 13, 2019, from https://www.google.com/search/howsearchworks/mission/

Green, J. T. (2017). *Pathways to creating Onkwehonwenéha speakers at Six Nations of the Grand River Territory*. https://www.snpolytechnic.com/researchpathways-creating-speakers-onkwehonwehneha-six-nations

Gupta, V., & Boulianne, G. (2020a). Automatic transcription challenges for Inuktitut, a low-resource polysynthetic language. *Proceedings of the 12th language resources and evaluation conference*, 2521–2527.

Gupta, V., & Boulianne, G. (2020b). Speech transcription challenges for resource constrained Indigenous language Cree. *Proceedings of the 1st joint workshop on spoken language technologies for under-resourced languages (SLTU) and collaboration and computing for under-resourced languages (CCURL)*, 362–367.

Hutton, C. (2010). Who owns language? Mother tongues as intellectual property and the conceptualization of human linguistic diversity. *Language Sciences*, *32*(6), 638–647. https://doi.org/10.1016/j.langsci.2010.06.001

Jimerson, R., & Prud'hommeaux, E. (2018). ASR for documenting acutely under-resourced Indigenous languages. In N. Calzolari, K. Choukri, C. Cieri, T. Declerck, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, S. Piperidis, T. Tokunaga, S. Goggi, & H. Mazo (Eds.), *Proceedings of the eleventh international conference on language resources and evaluation ([LREC]-2018)* (pp. 4161–4166). European Languages Resources Association (ELRA). https://www.aclweb.org/anthology/L18-1657

Jimerson, R., Simha, K., Ptucha, R., & Prudhommeaux, E. (2018). Improving ASR output for endangered language documentation. *The 6th intl. workshop on spoken language technologies for under-resourced languages*, 182–186. https://www.isca-speech.org/archive/SLTU_2018/abstracts/Robbie.html

Kazantseva, A., Maracle, B. O., Maracle, J. R., & Pine, A. (2018). Kawennón:nis: The wordmaker for Kanyen'kéha. In J. L. Klavans (Ed.), *Proceedings of the workshop on computational modeling of polysynthetic languages* (pp. 53–64). Association for Computational Linguistics. https://www.aclweb.org/anthology/W18-4806

Keegan, T. T. (2019, June 25). *Issues with Māori sovereignty over Māori language data* [Video File]. Available from http://video.web.gov.bc.ca/public/fpcc/letlanguageslive.html.

Keegan, T. T., & Cunliffe, D. (2014). Young people, technology and the future of te Reo Māori. In R. Higgins, R. Poia, & V. Olsen-Reeder (Eds.), *The value of the Māori language: Te Hua o te Reo Māori* (pp. 385–398). Huia Publishers.

Keegan, T. T., & Nfato, P. (2014). It's harder in my language but I still choose it. In P. Heinrich & N. Ostler (Eds.), *Proceedings of the 18th FEL conference on Indigneous languages: their value to the community* (pp. 43–48). https://hdl.handle.net/10289/10515

Keegan, T. T., & Sciascia, A. D. (2018). Hangarau me te Māori: Māori and technology. In M. Reilly, S. Duncan, G. Leoni, L. Paterson, L. Carter, M. Rātima, & P. Rewi (Eds.), *Te Kōparapara: An introduction to the Māori world* (pp. 359–371). Auckland University Press. https://hdl.handle.net/10289/11955

King, D. H. (1975). *A grammar and dictionary of the Cherokee language*. University of Georgia.

Kuhn, R., Davis, F., Désilets, A., Joanis, E., Kazantseva, A., Knowles, R., Littell, P., Lothian, D., Pine, A., Running Wolf, C., Santos, E., Stewart, D., Boulianne, G., Gupta, V., Maracle Owennatékha, B., Martin, A., Cox, C., Junker, M.-O., . . . Souter, H. (2020). The Indigenous languages technology project at NRC Canada: An empowerment-oriented approach to developing language software. *Proceedings of the 28th international conference on computational linguistics*, 5866–5878. https://doi.org/10.18653/v1/2020.coling-main.516

Lessard, G., Brinklow, N. T., & Levison, M. (2018). Natural language generation for polysynthetic languages: Language teaching and learning software for Kanyen'kéha (Mohawk). In J. L. Klavans (Ed.), *Proceedings of the workshop on computational modeling of polysynthetic languages* (pp. 41–52). Association for Computational Linguistics. https://www.aclweb.org/anthology/W18-4805

Littell, P., Kuhn, R., Hall, A., Kazantseva, A., Pine, A., & Cox, C. (2018). Indigenous language technologies in Canada: Assessment, challenges, and successes. In J. L. Klavans (Ed.), *Proceedings of the 27th international conference on computational linguistics* (pp. 2620–2632). https://www.aclweb.org/anthology/papers/C/C18/C18-1222/

Lothian, D., Akçayır, G., & Demmans Epp, C. (2019). Accommodating Indigenous people when using technology to learn their ancestral language. *Proceedings of the international workshop on supporting lifelong learning*. http://ceur-ws.org/Vol-2395/

Malone, G. (2017, April 5). First Nations take to designing their own apps to preserve their languages and cultures. *The Financial Post*. https://business.financialpost.com/entrepreneur/first-nations-take-to-designing-their-own-apps-to-preserve-their-languages-and-cultures

Mato, P., Keegan, T. T., & Naera, L. (2016). How usable is a smartphone with a Māori-language interface? *MAI Journal: A New Zealand Journal of Indigenous Scholarship*, *5*(1), 17–32. https://doi.org/10.20507/maijournal.2016.5.1.2

McIvor, O., & Anisman, A. (2018). Keeping our languages alive: Strategies for Indigenous language revitalization and maintenance. In Y. Watanabe (Ed.), *Handbook of cultural security* (pp. 90–109). Edward Elgar Publishing. https://doi.org/10.4337/9781786437747.00011

Micher, J. C. (2018). Using the Nunavut Hansard data for experiments in morphological analysis and machine translation. In J. L. Klavans (Ed.), *Proceedings of workshop on polysynthetic languages* (pp. 65–72). Association for Computational Linguistics. https://www.aclweb.org/anthology/W18-4807

Miller, B. (2020). Is technology value-neutral? *Science Technology and Human Values*, 1–28. https://doi.org/10.1177/0162243919900965

National Research Council of Canada. (n.d.). *Canadian Indigenous languages technology project*. https://nrc.canada.ca/en/research-development/research-collaboration/programs/canadian-indigenous-languages-technology-project

Outakoski, H., Cocq, C., & Steggo, P. (2018). Strengthening Indigenous languages in the digital age: social media-supported learning in Sápmi. *Media International Australia, 169*(1), 21–31. https://doi.org/10.1177/1329878X18803700

Parins, J. W. (2013). *Literacy and intellectual life in the Cherokee Nation, 1820–1906*. University of Oklahoma Press.

Pine, A., & Turin, M. (2018). Seeing the Heiltsuk orthography from font encoding through to Unicode: A case study using Convertextract. In C. Soria, L. Besacier, & L. Pretorius (Eds.), *Proceedings of the LREC 2018 workshop "CCURL 2018 – Sustaining knowledge diversity in the digital age"* (pp. 27–30). http://aidanpine.ca/static/cv/pdfs/Pine-Turin-Convertextract-2018.pdf

Pool, I. (2016). Colonialism's and postcolonialism's fellow traveller: The collection, use and misuse of data on Indigenous people. In T. Kukutai & J. Taylor (Eds.), *Indigenous data sovereignty* (Vol. 38, pp. 57–76). ANU Press. https://www.jstor.org/stable/j.ctt1q1crgf.11

Rice, K., & Thieberger, N. (2018). Tools and technology for language documentation and revitalization. In K. L. Rehg & L. Campbell (Eds.), *The Oxford handbook of endangered languages*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190610029.013.13

Schillo, J., & Turin, M. (2020). Applications and innovations in typeface design for North American Indigenous languages. *Book 2.0*, *10*(1), 71–98. https://doi.org/10.1386/btwo_00021_1

Sinclair, N. (2020, July 4). Learning Indigenous language gives community a future. *The Winnipeg Free Press*. https://www.winnipegfreepress.com/local/learning-indigenous-language-gives-community-a-future-571625612.html

Te Hiku Media. (2018). *Kaitiakitanga license*. https://github.com/TeHikuMedia/Kaitiakitanga-License/blob/master/LICENSE.md

The Mohawk Language Steering Committee. (1993). *The Mohawk language standardization project: Conference report August 9–10, 1993*. https://kanienkeha.net/the-mohawk-language-standardisation-project/

The Unicode Consortium. (n.d.). *Unicode–The world standard for text and emoji*. Retrieved December 14, 2019, from https://home.unicode.org/

Translation Commons. (2019). *Indigenous languages: Zero to digital*. https://sites.google.com/translationcommons.org/iyil/home/language-technology-development

Truth and Reconciliation Commission of Canada. (2015). *Truth and reconciliation commission of Canada: Calls to action*. Truth and Reconciliation Commission. http://trc.ca/assets/pdf/Calls_to_Action_English2.pdf

UNESCO Ad Hoc Expert Group on Endangered Languages. (2003). *Language vitality and endangerment*. http://www.unesco.org/new/fileadmin/MULTIMEDIA/HQ/CLT/pdf/Language_vitality_and_endangerment_EN.pdf

Wagner, I. (2017). New technologies, same ideologies: Learning from language revitalization online. *Language Documentation & Conservation*, *11*, 133–156. http://hdl.handle.net/10125/24730

Walter, M., & Suina, M. (2019). Indigenous data, Indigenous methodologies and Indigenous data sovereignty. *International Journal of Social Research Methodology*, *22*(3), 233–243. https://doi.org/10.1080/13645579.2018.1531228

Ward, M. (2004). The additional uses of CALL in the endangered language context. *ReCALL*, *16*(2), 345–359. https://doi.org/10.1017/S0958344004000722

Whaanga, H., Bainbridge, D., Anderson, M., Scrivener, K., Cader, P., Roa, T., & Keegan, T. T. (2015). He matapihi mā mua, mō muri: The ethics, processes, and procedures associated with the digitization of indigenous knowledge—the Pei Jones collection. *Cataloging and Classification Quarterly*, *53*(5–6), 520–547. https://doi.org/10.1080/01639374.2015.1009670

Wilks, J., Kennedy, G., Drew, N., & Wilson, K. (2018). Indigenous data sovereignty in higher education Towards a decolonised data quality framework. *Australian Universities' Review, 60*(2), 4–15.

YouTube About (n.d.). *YouTube for Press.* Retrieved November 24, 2019, from https://www.youtube.com/intl/en-GB/about/press/

Zahrer, A., Zgank, A., & Schuppler, B. (2020). Towards building an automatic transcription system for language documentation: Experiences from Muyu. *Proceedings of the 12th language resources and evaluation conference* (pp. 2893–2900). https://www.aclweb.org/anthology/2020.lrec-1.353/

Zhao, Y. (2003). Recent developments in technology and language learning: A literature review and meta-analysis. *CALICO Journal*, *21*(1), 7–27. https://doi.org/10.1558/cj.v21i1.7-27